IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

In re Patent Application of	)
	)
Monika Henzinger et al.	) **ATTN: APPEAL BRIEF - PATENTS**
	)
Application No.: 10/672,248	) Group Art Unit: 2176
	)
Filed: September 29, 2003	) Examiner: A. Rutledge
	)
For: IDENTIFICATION OF WEB	)
SITES THAT CONTAIN SESSION	)
IDENTIFIERS	)

U.S. Patent and Trademark Office
Customer Window, Mail Stop Appeal Brief - Patents
Randolph Building
401 Dulany Street
Alexandria, VA 22314

APPEAL BRIEF

This Appeal Brief is submitted in response to the Notice of Panel Decision from Pre-Appeal Brief Review, dated July 18, 2007, and in support of the Notice of Appeal, filed May 18, 2007.

I. <u>REAL PARTY IN INTEREST</u>

The real party in interest in this appeal is Google Inc.

II. <u>RELATED APPEALS, INTERFERENCES, AND JUDICIAL PROCEEDINGS</u>

Appellants are unaware of any related appeals, interferences, or judicial proceedings.

III.   STATUS OF CLAIMS

Claims 1-8, 10, 12-21, 23-26, and 28-30 are pending in this application.

Claims 1-8, 10, 12-21, 23-26, and 28-30 stand rejected under 35 U.S.C. § 103(a) as being

unpatentable over the PCT International Publication Number WO 03/017023 to Galai et al.

("Galai") in view of U.S. Patent Number 6,665,658 to DaCosta et al. ("DaCosta").

Claims 9, 11, 22, and 27 were previously canceled without prejudice or disclaimer.

Claims 1-8, 10, 12-21, 23-26, and 28-30 are the subject of the present appeal. These

claims are reproduced in the Claim Appendix of this Appeal Brief.

IV.   STATUS OF AMENDMENTS

No claim amendments were filed subsequent to the non-final Office Action, dated

February 7, 2007.

A Pre-Appeal Request for Review was filed on May 18, 2007. A Notice of Panel

Decision from Pre-Appeal Brief Review was issued on July 18, 2007 indicating that the

application remains under appeal.

V.   SUMMARY OF CLAIMED SUBJECT MATTER

In the paragraphs that follow, a concise explanation of the independent claims and the

claims reciting means-plus-function or step-plus-function language that are involved in this

appeal will be provided by referring, in parenthesis, to examples of where support can be found

in the specification and drawings.

Claim 1 recites a method for crawling documents. The method includes receiving a uniform resource locator (URL) [Fig. 6, act 601; and page 13, line 17 – page 14, line 1]; receiving at least two different copies of a document associated with the URL [Fig. 6, act 602; and p. 14, lines 3 – 10]; and determining whether a web site corresponding to the URL uses session identifiers based on a comparison of URLs that are within the document and that change between the at least two different copies of the document, where the web site is determined to use session identifiers when a portion of the URLs that change between the at least two different copies of the document is greater than a threshold [Fig. 6, acts 604 – 607; and page 15, lines 9 – 19].

Claim 4 further defines the method of claim 1 and recites that the compared URLs that change include URLs that are local to the web site [Fig. 6, act 603; and p. 14, line 18 – page 15, line 2].

Claim 10 is directed to a method for identifying web sites that use session identifiers. The method of claim 10 includes downloading at least two different copies of at least one document from a web site [Fig. 6, act 602; and page 14, lines 3 – 10]; extracting uniform resource locators (URLs) from the two different copies of the web document [Fig. 6, act 603; and page 14, line 18 – page 15, line 2]; comparing the extracted URLs of the two different copies of the document [Fig. 6, act 604; and page 15, lines 11 – 14]; and determining whether the web site uses session identifiers when the comparison indicates that at least a portion of the URLs change between the two different copies [Fig. 6, acts 605 and 606; and page 15, lines 15 – 19].

Claim 12 further defines the features of claim 10 and recites that extracting URLs from the two different copies of the document includes extracting only URLs that are local to the web

site [Fig. 6, act 603; and p. 14, line 18 – page 15, line 2].

Claim 15 is directed to a device that includes a spider component configured to crawl web documents associated with at least one web site [Fig. 3, element 315; and page 10, lines 8 - 21]. The device additionally includes a session identifier component configured to determine whether the web site uses session identifiers based on a comparison of a portion of uniform resource locators (URLs) that change between different copies of at least one web document downloaded from the web site [Fig. 3, element 320; Fig. 6, acts 604 – 607; page 10, line 27 – page 11, line 4; and page 15, lines 9 – 19].

Claim 19 further defines the method of claim 15 and recites that the portion of the URLs that change are identified from URLs that are local to the web site [Fig. 6, act 603; and p. 14, line 18 – page 15, line 2].

Claim 21 is directed to device that includes means for downloading at least two different copies of at least one web document from a web site [Fig. 6, act 602; and page 14, lines 3 – 10]. The device additionally includes means for extracting uniform resource locators (URLs) from the two different copies of the web document [Fig. 6, act 603; and page 14, line 18 – page 15, line 2]; means for comparing the extracted URLs of the two different copies of the web document [Fig. 6, act 604; and page 15, lines 11 – 14]; and means for determining whether the web site uses session identifiers when the comparison indicates that at least a portion of the URLs change between the two different copies [Fig. 6, acts 605 and 606; and page 15, lines 15 – 19].

Claim 26 is directed to a computer-readable medium containing programming instructions [p. 9, lines 1-9; p. 17, line 16 to p. 18, line 3] that when executed by at least one processor cause the processor to perform a method for identifying web sites that use session

identifiers. The method includes downloading at least two different copies of at least one

document from a web site [Fig. 6, act 602; and page 14, lines 3 – 10]; extracting uniform

resource locators (URLs) from the two different copies of the document [Fig. 6, act 603; and

page 14, line 18 – page 15, line 2]; comparing the extracted URLs of the two different copies of

the web document [Fig. 6, act 604; and page 15, lines 11 – 14]; and determining whether the web

site uses session identifiers when the comparison indicates that at least a portion of the URLs

change between the two different copies [Fig. 6, acts 605 and 606; and page 15, lines 15 – 19].

VI.   GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

A.   Claims 1-8, 10, 12-21, 23-26, and 28-30 stand rejected under 35 U.S.C. § 103(a)

as unpatentable over Galai and DaCosta.

VII.   ARGUMENT

**A.   The Rejection Under 35 U.S.C. § 103(a) Based on Galai
and DaCosta Should be Reversed.**

The initial burden of establishing a prima facie basis to deny patentability to a claimed

invention is always upon the Examiner. In re Oetiker, 977 F.2d 1443, 24 USPQ2d 1443 (Fed.

Cir. 1992). In rejecting a claim under 35 U.S.C. § 103, the Examiner must provide a factual

basis to support the conclusion of obviousness. In re Warner, 379 F.2d 1011, 154 USPQ 173

(CCPA 1967). Based upon the objective evidence of record, the Examiner is required to make

the factual inquiries mandated by Graham v. John Deere Co., 86 S.Ct. 684, 383 U.S. 1, 148

USPQ 459 (1966). KSR International Co. v. Teleflex Inc., 550 U.S. 127 S. Ct. 1727 (2007).

The Examiner is also required to explain how and why one having ordinary skill in the art would

have been led to modify an applied reference and/or combine applied references to arrive at the

claimed invention. Uniroyal, Inc. v. Rudkin-Wiley Corp., 837 F.2d 1044, 5 USPQ2d 1434 (Fed.

Cir. 1988).

1.   Claims 1-3 and 5-8.

Each of the independent claims, including independent claim 1, recites, among other

things, determining whether a web site uses "session identifiers." A session identifier, as is

known in the art and is consistently used in the pending specification, refers to embedded

information within the URL of a web page. (See Spec., for example, paragraphs 0006, 0042, and

0043; and Fig. 5). Session identifiers are commonly used by web sites to track user behavior as

they traverse a web site. Fig. 5 of the specification illustrates examples of the presence of session

identifiers in URLS. (Spec., paragraph 0042 and Fig. 5).

Claim 1 is directed to a method for crawling documents that includes receiving a uniform

resource locator (URL) and receiving at least two different copies of a document associated with

the URL. The method further includes determining whether a web site corresponding to the URL

uses session identifiers based on a comparison of URLs that are within the document and that

change between the at least two different copies of the document, where the web site is

determined to use session identifiers when a portion of the URLs that change between the at least

two different copies of the document is greater than a threshold.

Neither Galai nor DaCosta, whether taken alone or in any reasonable combination,

discloses or suggests the combination of features recited in claim 1. For example, Galai and

DaCosta do not disclose or suggest determining, in the manner recited in claim 1, whether a web

site corresponding to a URL uses session identifiers. More specifically, neither Galai nor

DaCosta, discloses or suggests determining whether a web site corresponding to the URL uses

session identifiers based on a comparison of URLs that are within the document and that change

between the at least two different copies of the document, where the web site is determined to

use session identifiers when a portion of the URLs that change between the at least two different

copies of the document is greater than a threshold.

The Examiner alleges that Galai discloses this feature and cites various portions of pages

20, 21, 27, and 28 of Galai for support. (Final Office Action, page 3.) Appellants disagree with

the Examiner's interpretation of Galai.

Galai is directed to "a system and a method for automatically extracting content from a

document such as a Web page, and for submitting such content to a search engine." (Galai, page

4, lines 3-5.) At pages 20 and 21, Galai discusses a technique for "normalizing the URI of the

document, such as the URL of a web page … in order to index substantially similar Web pages

only once." (Galai, page 20, lines 10-14.) Galai recognizes that Web pages that use session IDs

may be similar although not completely identical. (Galai, page 20, lines 21-23.)

Although Galai discloses the existence of session identifiers, Galai does not determine

whether a web site uses session identifiers using the specific technique recited in claim 1. That

is, Galai does not disclose or suggest, as recited in claim 1, comparing URLs that are within a

document and that change between the at least two different copies of the document, where the

web site is determined to use session identifiers when a portion of the URLs that change between

the at least two different copies of the document is greater than a threshold.

In contrast to this feature of claim 1, and as acknowledged by the Examiner, Galai appears to disclose comparing a web page with a second web page, which was accessed with a reduced version of the URL used to access the first web page, to determine if the two web pages are similar. (Final Office Action, page 3 and Galai, page 20, lines 13-20). That is, Galai discloses directly comparing web pages to determine if the web pages are similar. Galai notes that if the two web pages are similar, this may indicate that the parameter (i.e., a divisible subunit of a URL) used to reduce the URL that was used to access the two web pages is redundant. (Galai, page 20, lines 21 and 22). Galai, however, clearly discloses that that procedure is based on a comparison of the two web pages. Galai discloses additional details about the techniques used for comparing web pages at page 21. (See Galai, page 21, lines 5-20). For example, Galai discloses that the web page comparison function may be based on a comparison for similarity in content or a comparison for visual similarity.

Comparing a web page for similarity in content or visual similarity, as disclosed by Galai, however, cannot be said to disclose or suggest determining whether a web site corresponding to a URL uses session identifiers "based on a comparison of URLs that are within the document and that change between the at least two different copies of the document, where the web site is determined to use session identifiers when a portion of the URLs that change between the at least two different copies of the document is greater than a threshold," as recited in claim 1 (emphasis added). A comparison function that "checks for similarity in content and more preferably produces a similarity level, which is the likelihood of the two Web pages to have the same content," as described by Galai at page 21, lines 4-7, does not disclose or suggest the

determination recited in claim 1. If anything, Galai's explicit disclosure of checking for

similarity of Web pages based on the <u>content</u> of the web page teaches away from the

determination made in claim 1, which is based on <u>a portion of URLs</u> that change.

The Examiner additionally points to portions of pages 27 and 28 of Galai as allegedly

disclosing "comparing a portion of the URLs that change between the two copies of the

document and determining a similarity based on a predetermined value of the portion of the

URLs that change." (Final Office Action, page 3.) Pages 27 and 28 of Galai disclose material

similar to that disclosed at pages 20 and 21 of Galai. Specifically, Galai discloses comparing the

<u>content</u> of a web page to determine similarity or comparing visual layout characteristics to

determine similarity. (Galai, page 27, line 14 through page 28, line 21.) These sections of Galai,

however, do not disclose or suggest making a determination "when a portion of the URLs that

change between the at least two different copies of the document is greater than a threshold," as

recited in claim 1. Comparing the "content" of documents, as described by Galai, simply does

not disclose or suggest this feature of claim 1.

In the Final Office Action, the Examiner appears to contend that a comparison of two

web pages, as disclosed by Galai, discloses the features recited in claim 1, because "claim 1 also

recites a comparison of links *within two web pages*." (Final Office Action, pages 12 and 13,

numbered paragraph 5 in "Response to Arguments" section, emphasis in original.) Appellants

submit that the Examiner appears to be misreading the features of claim 1. Although claim 1

does recite "comparison of URLs" and "at least two different copies of the document," claim 1

recites more than simply the comparison of URLs within documents. That is, the determination

of claim 1 is based on a comparison of URLs that are within the document and that change

between the at least two different copies of the document, where the web site is determined to

use session identifiers when a portion of the URLs that change between the at least two different

copies of the document is greater than a threshold. Comparing entire web pages (that contain

links) for similarity, as disclosed by Galai, cannot be said to suggest this feature of claim 1.

The Examiner further appears to contend that page 27, lines 14-16 of Galai is particularly

relevant to the features recited in claim 1. (Final Office Action, page 13.) This section of Galai,

which is identical to page 20, lines 21-23 of Galai, states:

> If the parameter is redundant, the Web pages may be expected to be similar,
> although perhaps not completely identical. Lack of identity may occur if the Web
> page includes one or more links with the complete URL, as for a session ID.

This section of Galai recognizes that web pages that use session IDs may be similar, although not

identical. In the same paragraph, Galai goes on to further state:

> For that reason, the comparison function of the present invention preferably
> checks for similarity in content and more preferably produces a similarity level,
> which is the likelihood of the two Web pages to have the same content. If this
> value exceeds a certain threshold, then most preferably the removed parameter is
> considered to be redundant.

(Galai, page 27, lines 18-22.) Here, Galai discloses checking web pages for similarity in content

using a comparison function that produces a similarity level. In this paragraph and in succeeding

paragraphs, Galai makes it clear that the comparison function is based on a comparison of the

content of a web page. Neither the comparison function nor any other element of Galai,

however, in any way discloses or suggests, as recited in claim 1, determining that a web site uses

session identifiers when a portion of the URLs that change between the at least two different

copies of a document is greater than a threshold.

The Examiner appears to rely on DaCosta for "the specific purpose of determining

whether the web site uses session identifiers." (Final Office Action, page 4.) DaCosta, however,

does not even mention session identifiers, much less determining whether a web site uses session

identifiers.

The Examiner points to portions of columns 4, 5, and 6 of DaCosta as disclosing the use

of session identifiers. Specifically, the Examiner appears to contend that DaCosta, at column 4,

lines 41 through column 5, line 23 and column 6, lines 21-40, discloses session identifiers.

(Office Action, page 3). Appellants respectfully disagree with the Examiner's interpretation of

DaCosta.

Column 4, line 31 through column 5, line 23 of DaCosta discusses, among other things,

"session data" of a web site. For example, DaCosta states:

> It is also preferred that the step of determining if said URL is a dynamic website
> further comprise performing a hypertext transfer protocol GET method of the
> website, downloading a content including a header of the website, and scanning
> the header for the session data which may be represented by a cookie.

(DaCosta, column 4, lines 40-46) (emphasis added). The "session data" discussed in this section

of DaCosta appears to broadly relate to any session information used in the context of dynamic

generation of web content. DaCosta discloses that the session data may be obtained by scanning

a header of a website, and that the session data may be represented by a cookie. The Examiner

can appreciate session data represented by a cookie, as disclosed by DaCosta, cannot be said to

be equivalent to the session identifier recited in claim 1. As described in the pending

specification, a cookie, although may be used to track user behavior, is different than a session

identifier. (Spec., first sentence of paragraph 0006).

The Examiner also points to column 6, lines 21-40 of DaCosta as disclosing the session

identifiers recited in claim 1. This section of DaCosta, however, again discusses "session data" that may be represented in a cookie.

DaCosta, as discussed above, does not disclose using session identifiers. DaCosta therefore, cannot be relied upon to suggest, as alleged by the Examiner, "that the URLs [of Galai] are compared for the specific purpose of determining whether the web site uses session identifiers." (Final Office Action, page 4.) Thus, DaCosta does not cure the above-noted deficiencies of Galai.

Accordingly, Galai and DaCosta, even if combined as the Examiner suggests, would not disclose or suggest each of the features recited in claim 1. For at least these reasons, it is respectfully submitted that claims 1-3 and 5-8 are patentable over Galai and DaCosta, whether taken alone or in any reasonable combination, under 35 U.S.C. § 103. Reversal of the rejection of claims 1-3 and 5-8 is respectfully requested.

      2.   Claim 4.

Dependent claim 4 recites that the compared URLs that change include URLs that are local to the web site.

Initially, claim 4 depends from claim 3, which depends from claim 1. Claim 4 is, therefore, patentable over Galai and DaCosta, whether taken alone or in any reasonable combination, for at least the reasons given with regard to claim 1.

Further, Galai and DaCosta, whether taken alone or in any reasonable combination, do not disclose or suggest the features recited in claim 4.

Regarding claim 4, the Examiner states:

> Galai teaches that the method of comparing URLs can be applied to any web page in a site (p. 4, l. 15-20). Galai teaches an automatic method of URL comparison to remove redundant parameters from pages (p. 20, l. 10-20), which would include session IDs (p. 20, l. 21-23), where the rules are determined automatically by comparing the URLs for redundancy and normalizing them.

(Final Office Action, page 5.) Initially, and as discussed previously with respect to claim 1, Appellants note that, contrary to the Examiner's statements, Galai does not teach an automatic method of URL comparison. Instead, Galai discloses comparing entire web pages to determine similarity in content between web pages. (Galai, pages 20 and 21.) Therefore, Galai could not disclose or suggest, as recited in claim 4, that the compared URLs that change include URLs that are local to the web site.

Further, Galai does not discuss any particular usage of URLs local to a web site. For this reason also, Galai could not disclose or suggest, as recited in claim 4, that the compared URLs that change include URLs that are local to the web site.

In rejecting claim 4, the Examiner also cited page 4, lines 15-20 of Galai as disclosing that the method of Galai "can be applied to any web page within a site." (Final Office Action, page 5.) Page 4, lines 15-20 of Galai merely discloses a list of document types for which a search engine may process URIs. This section of Galai, however, cannot be said to disclose or suggest that the compared URIs are URIs that are local to a web site.

DaCosta does not cure the above-noted deficiencies of Galai. Accordingly, Galai and DaCosta do not disclose or suggest each of the features recited in claim 4. For at least these reasons, it is respectfully submitted that claim 4 is patentable over Galai and DaCosta, whether taken alone or in any reasonable combination, under 35 U.S.C. § 103. Reversal of the rejection of claim 4 is respectfully requested.

3.   Claims 10, 13, 14, 21, 24, and 25.

Independent claim 10 is directed to a method for identifying web sites that use session identifiers. The method includes downloading at least two different copies of at least one document from a web site; extracting uniform resource locators (URLs) from the two different copies of the web document; comparing the extracted URLs of the two different copies of the document; and determining whether the web site uses session identifiers when the comparison indicates that at least a portion of the URLs change between the two different copies.

Neither Galai nor DaCosta, whether taken alone or in any reasonable combination, discloses or suggests the combination of features recited in claim 10. For example, Galai and DaCosta do not disclose or suggest comparing extracted URLs of two different copies of a document and determining whether a web site uses session identifiers when a comparison indicates that at least a portion of the URLs change between the two different copies.

As discussed above with respect to claim 1, Galai notes that if two web pages are similar, this may indicate that a parameter used to reduce the URL through which the second web page was obtained is redundant. (Galai, page 20, lines 21 and 22). Galai clearly discloses determining whether the parameter used to reduce the URL is redundant based on a comparison of the two web pages. Comparing web pages for similarity in content or visual similarity, as described by Galai, does not disclose or suggest the features of claim 10, which include comparing the extracted URLs of the two different copies of the document and determining whether the web site uses session identifiers when the comparison indicates that at least a portion of the URLs change between the two different copies. Galai does not in any way use information relating to a

- 14 -

portion of URLs that change between web pages, and thus could not disclose or suggest, as recited in claim 10, a comparison that indicates that at least a portion of the URLs change between the two different copies of the document.

The Examiner alleges that Galai discloses many of the features of claim 10 and cites various portions of pages 20, 21, 27, and 28 of Galai for support. (Final Office Action, pages 6 and 7.) These sections of Galai were discussed previously with respect to claim 1. Appellants submit that neither these section of Galai, nor any other section of Galai, discloses or suggests, as recited in claim 10, comparing the extracted URLs of the two different copies of the document and determining whether the web site uses session identifiers when the comparison indicates that at least a portion of the URLs change between the two different copies.

The Examiner relies on DaCosta for "the specific purpose of determining whether the web site uses session identifiers." (Final Office Action, page 7.) As discussed previously with respect to claim 1, however, DaCosta does not even mention session identifiers, much less determining whether a web site uses session identifiers. Thus, DaCosta cannot cure the above-noted deficiencies of Galai.

Accordingly, Galai and DaCosta, even if combined as the Examiner suggests, would not disclose or suggest each of the features recited in claim 10. For at least these reasons, it is respectfully submitted that claims 10, 13, 14, 21, 24, and 25 are patentable over Galai and DaCosta, whether taken alone or in any reasonable combination, under 35 U.S.C. § 103. Reversal of the rejection of claims 10, 13, 14, 21, 24, and 25 is respectfully requested.

    4.   Claims 12, 23, and 28.

Dependent claim 12 recites that extracting URLs from the two different copies of the document includes extracting only URLs that are local to the web site.

Initially, claim 12 depends from claim 10. Claim 12 is, therefore, patentable over Galai and DaCosta, whether taken alone or in any reasonable combination, for at least the reasons given with regard to claim 10.

Further, Galai and DaCosta, whether taken alone or in any reasonable combination, do not disclose or suggest the features recited in claim 12.

Regarding claim 12, the Examiner states: "Galai teaches that the method of comparing URLs can be applied to any web page in a site (p. 4, l. 15-20)." (Final Office Action, page 8.) Initially, and as discussed previously, Appellants note that, contrary to the Examiner's statements, Galai does not teach any specific technique for comparing URLs. Instead, Galai discloses comparing entire web pages to determine similarity in content between web pages. (Galai, pages 20 and 21.)

In any event, Appellants submit that Galai does not disclose or suggest extracting URLs in the manner recited in claim 12. Galai does not particularly mention, in any manner, using URLs local to a web site, much less where extracting URLs from two different copies of a document includes extracting only URLs that are local to a web site.

DaCosta does not cure the above-noted deficiencies of Galai. Accordingly, Galai and DaCosta do not disclose or suggest each of the features recited in claim 12. For at least these reasons, it is respectfully submitted that claims 12, 23, and 28 are patentable over Galai and DaCosta, whether taken alone or in any reasonable combination, under 35 U.S.C. § 103. Reversal of the rejection of claims 12, 23, and 28 is respectfully requested.

     5.   Claims 15-18 and 20.

Independent claim 15 is directed to a device comprising a spider component and a session identifier component. The spider component is configured to crawl web documents associated with at least one web site. The session identifier component is configured to determine whether the web site uses session identifiers based on a comparison of a portion of the URLs that change between different copies of at least one web document downloaded from the web site.

Neither Galai nor DaCosta, whether taken alone or in any reasonable combination, disclose or suggest the combination of features recited in claim 15. For example, Galai and DaCosta do not disclose or suggest a session identifier component configured to determine whether the web site uses session identifiers based on a comparison of a portion of the URLs that change between different copies of at least one web document downloaded from the web site.

As discussed above with respect to claim 1, Galai notes that if the two web pages are similar, this may indicate that a parameter used to reduce the URL through which the second web page was obtained is redundant. (Galai, page 20, lines 21 and 22). Galai clearly discloses determining whether the parameter used to reduce the URL is redundant based on a comparison of the two <u>web pages</u>. Comparing web pages for similarity in content or visual similarity, as described by Galai, does not disclose or suggest the session identifier component recited in claim 15. That is, Galai does not disclose or suggest a session identifier component configured to determine whether a web site uses session identifiers based on <u>a comparison of a portion of the URLs that change</u> between different copies of at least one web document downloaded from the web site.

The Examiner alleges that Galai discloses many of the features of claim 15 and cites various portions of pages 20, 21, 27, and 28 of Galai for support. (Final Office Action, page 9.) These sections of Galai were discussed previously with respect to claim 1. Appellants submit that neither these sections of Galai, nor any other section of Galai, discloses or suggests, as recited in claim 15, a session identifier component configured to determine whether a web site uses session identifiers based on a comparison of a portion of the URLs that change between different copies of at least one web document downloaded from the web site.

The Examiner relies on DaCosta for "the specific purpose of determining whether the web site uses session identifiers." (Final Office Action, pages 9 and 10.) As discussed previously with respect to claim 1, however, DaCosta does not even mention session identifiers, much less determining whether a web site uses session identifiers. Thus, DaCosta cannot cure the above-noted deficiencies of Galai.

Accordingly, Galai and DaCosta, even if combined as the Examiner suggests, would not disclose or suggest each of the features recited in claim 15. For at least these reasons, it is respectfully submitted that claims 15-18 and 20 are patentable over Galai and DaCosta, whether taken alone or in any reasonable combination, under 35 U.S.C. § 103. Reversal of the rejection of claims 15-18 and 20 is respectfully requested.

> 6.   Claim 19.

Dependent claim 19 recites that the portion of the URLs that change are identified from URLs that are local to the web site.

Initially, claim 19 depends from claim 15. Claim 19 is, therefore, patentable over Galai

and DaCosta, whether taken alone or in any reasonable combination, for at least the reasons given with regard to claim 15.

Further, Galai and DaCosta, whether taken alone or in any reasonable combination, do not disclose or suggest the combination of features recited in claim 19.

Regarding claim 19, the Examiner states: "Galai teaches that the method of comparing URLs can be applied to any web page in a site (p. 4, l. 15-20)." (Final Office Action, page 11.) Initially, and as discussed previously, Appellants note that, contrary to the Examiner's statements, Galai does not teach any specific technique for comparing URLs. Instead, Galai discloses comparing entire web pages to determine similarity in content between web pages. (Galai, pages 20 and 21.)

Further, Galai does not discuss using, in any particular manner, URLs local to a web site. Thus, Galai could not disclose or suggest, as recited in claim 19, that the portion of the URLs that change are identified from URLs that are local to the web site.

DaCosta does not cure the above-noted deficiencies of Galai. Accordingly, Galai and DaCosta do not disclose or suggest each of the features recited in claim 19. For at least these reasons, it is respectfully submitted that claim 19 is patentable over Galai and DaCosta, whether taken alone or in any reasonable combination, under 35 U.S.C. § 103. Reversal of the rejection of claim 19 is respectfully requested.

VIII.   CONCLUSION

In view of the foregoing arguments, Appellants respectfully solicits the Honorable Board to reverse the Examiner's rejections of claims 1-8, 10, 12-21, 23-26, and 28-30 under 35 U.S.C. § 103.

To the extent necessary, a petition for an extension of time under 37 C.F.R. § 1.136 is hereby made. Please charge any shortage in fees due in connection with the filing of this paper, including extension of time fees, to Deposit Account No. 50-1070 and please credit any excess fees to such deposit account.

Respectfully submitted,

HARRITY SNYDER, L.L.P.

/Viktor Simkovic, Reg. No. 56012/
Viktor Simkovic
Reg. No. 56,012

Date: October 26, 2007
11350 Random Hills Road
Suite 600
Fairfax, Virginia 22030
(571) 432-0800

IX.   CLAIM APPENDIX

1.   A method for crawling documents comprising:

receiving a uniform resource locator (URL);

receiving at least two different copies of a document associated with the URL; and

determining whether a web site corresponding to the URL uses session identifiers based

on a comparison of URLs that are within the document and that change between the at least two

different copies of the document, where the web site is determined to use session identifiers

when a portion of the URLs that change between the at least two different copies of the

document is greater than a threshold.

2.   The method of claim 1, wherein the document is a home page of the web site.

3.   The method of claim 1, further comprising:

extracting, when the URL corresponds to a web site that uses session identifiers, a session

identifier from the URL to obtain a clean URL; and

determining whether the URL has already been crawled based on a comparison of the

clean URL to a set of clean URLs that represent previously crawled URLs.

4.   The method of claim 3, wherein the compared URLs that change include URLs

that are local to the web site.

5.   The method of claim 3, wherein the session identifiers from the URLs are

extracted using rules for the web site.

6.   The method of claim 5, wherein the rules are determined automatically.

7.   The method of claim 3, further comprising:

receiving the URL as a URL from a previously crawled web document.

8.   The method of claim 3, further comprising:

crawling the URL when the URL is determined to not already have been crawled.

9.   (canceled)

10.   A method for identifying web sites that use session identifiers comprising:

downloading at least two different copies of at least one document from a web site;

extracting uniform resource locators (URLs) from the two different copies of the web document;

comparing the extracted URLs of the two different copies of the document; and

determining whether the web site uses session identifiers when the comparison indicates that at least a portion of the URLs change between the two different copies.

11.   (canceled)

12.   The method of claim 10, wherein extracting URLs from the two different copies

of the document includes extracting only URLs that are local to the web site.

13.   The method of claim 10, wherein the document is a home page of the web site.

14.   The method of claim 10, further comprising:

analyzing the extracted URLs, when the web site is determined to use session identifiers,

to generate at least one rule identifying how the session identifiers are embedded in the URLs.

15.   A device comprising:

a spider component configured to crawl web documents associated with at least one web

site; and

a session identifier component configured to determine whether the web site uses session

identifiers based on a comparison of a portion of uniform resource locators (URLs) that change

between different copies of at least one web document downloaded from the web site.

16.   The device of claim 15, wherein the spider component further comprises:

at least one fetch component configured to download content from a network; and

a content manager configured to extract URLs from the downloaded content.

17.   The device of claim 16, wherein the spider component further comprises:

a URL manager configured to store the extracted URLs.

18.   The device of claim 15, wherein the at least one web document is a home page of the web site.

19.   The device of claim 15, wherein the portion of the URLs that change are identified from URLs that are local to the web site.

20.   The device of claim 15, further comprising:

a session rule generator configured to generate rules describing how the web site embeds session identifiers in the at least one web document.

21.   A device comprising:

means for downloading at least two different copies of at least one web document from a web site;

means for extracting uniform resource locators (URLs) from the two different copies of the web document;

means for comparing the extracted URLs of the two different copies of the web document; and

means for determining whether the web site uses session identifiers when the comparison indicates that at least a portion of the URLs change between the two different copies.

22.   (canceled)

23.   The device of claim 21, wherein the means for extracting URLs from the two

different copies of the web document includes means for extracting only URLs that are local to

the web site.

24.   The device of claim 21, wherein the web document is a home page of the web

site.

25.   The device of claim 21, further comprising:

means for analyzing the extracted URLs, when the web site is determined to use session

identifiers, to generate rules describing how the session identifiers are embedded in the URLs.

26.   A computer-readable medium containing programming instructions that when

executed by at least one processor cause the processor to perform a method for identifying web

sites that use session identifiers including:

downloading at least two different copies of at least one document from a web site;

extracting uniform resource locators (URLs) from the two different copies of the

document;

comparing the extracted URLs of the two different copies of the web document; and

determining whether the web site uses session identifiers when the comparison indicates

that at least a portion of the URLs change between the two different copies.

27.   (canceled)

28.   The computer-readable medium of claim 26, wherein extracting URLs from the two different copies of the web document includes extracting only URLs that are local to the web site.

29.   The computer-readable medium of claim 26, wherein the web document is a home page of the web site.

30.   The computer-readable medium of claim 26, further comprising instructions that cause the at least one processor to:

analyze the extracted URLs, when the web site is determined to use session identifiers, to generate at least one rule describing how the session identifiers are embedded in the URLs.

31.   (canceled)

X.   <u>EVIDENCE APPENDIX</u>

None

XI.   RELATED PROCEEDINGS APPENDIX

None